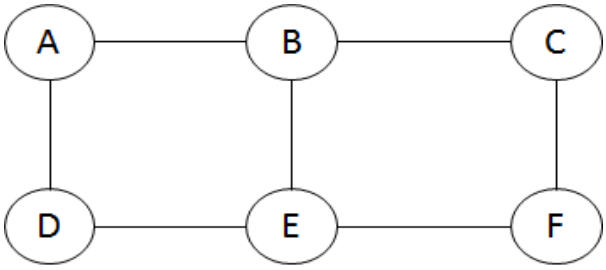
SVKM's
D. J. Sanghvi College of Engineering

Program: B.Tech in Mechanical Engineering **Academic Year: 2022** **Duration: 3 hours**
Date: 06.01.2023
Time: 10:30 am to 01:30 pm
Subject: Big Data Analysis (Semester VII) **Marks: 75**

Instructions: Candidates should read carefully the instructions printed on the question paper and on the cover page of the Answer Book, which is provided for their use.
(1) This question paper contains two pages.
(2) All Questions are Compulsory.
(3) All questions carry equal marks.
(4) Answer to each new question is to be started on a fresh page.
(5) Figures in the brackets on the right indicate full marks.
(6) Assume suitable data wherever required, but justify it.
(7) Draw the neat labelled diagrams, wherever necessary.

Question No.		Max. Marks
Q1 (a)	Explain different types of big data analytics with suitable examples. **OR** Discuss the advantages and limitations of Hadoop.	[05] [05]
Q1 (b)	What is MongoDB? State the features of MongoDB.	[10]
Q2 (a)	Explain the concept of replication while choosing distribution models. **OR** Compare and contrast MongoDB with RDBMS.	[05] [05]
Q2 (b)	Explain Hadoop Ecosystem in detail with a neat diagram.	[10]
Q3 (a)	Explain the following Big Data Technologies: i. Hadoop ii. MongoDB iii. Rapid Miner iv. Blockchain v. Tableau	[10]
Q3 (b)	Apply MapReduce on the following document to count the frequency of words. Show all the phases properly. foo car bar foo aar foo car bar bar	[05]
Q4 (a)	Explain the Content-based Recommendation System in detail with its advantages and disadvantages. **OR**	[05] [05]

	Define the terms Centrality, Degree Centrality, Closeness Centrality, Betweenness Centrality and Clique.	
Q4 (b)	Consider the user-item rating matrix as shown in the table below. There are five data ratings from 1 to 5 {1, 2, 3, 4, 5}. Predict the rating for "User1" on "Item5" using Pearson's Correlation approach of Collaborative filtering. Items/ Users | Item1 | Item 2 | Item 3 | Item 4 | Item 5 User1 | 1 | 3 | 3 | 2 | ? User2 | 2 | 4 | 2 | 2 | 4 User3 | 1 | 3 | 3 | 5 | 1 User4 | 1 | 5 | 3 | 3 | 2 User5 | 1 | 1 | 5 | 2 | 1 **OR** Explain Girvan-Newman Algorithm for Clustering. Using the same, find the communities in the following web graph. 	[10] [10]
Q5 (a)	List the functions of YARN in HDFS architecture.	[05]
Q5 (b)	Explain Arithmetic and Relational Operators in R with suitable examples. **OR** Write a R program 1. To create a bar plot of five subject marks. 2. To compute the sum, mean, and product of a given vector elements.	[10] [10]